

# Preferential Nucleosome Occupancy at High Values of DNA Helical Rise

FRANCESCO Pedone<sup>1</sup> and DANIELE Santoni<sup>2,3,\*</sup>

Department of Biology and Biotechnology, 'Sapienza' University, P.le A. Moro 5, 00185 Rome, Italy<sup>1</sup>; Barcelona Institute for Research in Biomedicine (IRB), Barcelona Science Park, C/Samitier 1-5, 08015 Barcelona, Spain<sup>2</sup> and Istituto di Analisi dei Sistemi e Informatica Antonio Ruberti, Consiglio Nazionale delle Ricerche, Viale Manzoni 30, 00185 Rome, Italy<sup>3</sup>

\*To whom correspondence should be addressed. E-mail: francesco.pedone@uniroma1.it (F. Pedone); santoni@caspur.it (D. Santoni).

(Received 15 July 2011; accepted 29 November 2011)

## Abstract

**Nucleosomes are the basic structural units of eukaryotic chromatin and play a key role in the regulation of gene expression. Nucleosome formation depends on several factors, including properties of the sequence itself, but also physical constraints and epigenetic factors such as chromatin-remodelling enzymes. In this view, a sequence-dependent approach is able to capture a general tendency of a region to bind a histone octamer. A reference data set of positioned nucleosomes of *Saccharomyces cerevisiae* was used to study the role of DNA helical rise in histone–DNA interaction. Genomic sequences were transformed into arrays of helical rise values by a tetranucleotide code and then turned into profiles of mean helical rise values. These profiles resemble maps of nucleosome occupancy, suggesting that intrinsic histone–DNA interactions are linked to helical rise. The obtained results show that preferential nucleosome occupancy occurs where the mean helical rise reaches its largest values. Mean helical rise profiles obtained by using maps of positioned nucleosomes of the *Drosophila melanogaster* and *Plasmodium falciparum* genomes, as well as *Homo sapiens* chromosome 20 confirm that nucleosomes are mainly located where the mean helical rise reaches its largest values.**

**Key words:** nucleosome; tetranucleotide; helical rise

## 1. Introduction

The observation that specific DNA sequences favour the formation of nucleosomes<sup>1–6</sup> raises the possibility that sequence plays a significant role in organizing nucleosomal arrays.<sup>7,8</sup> Recently, computational models for sequence-based prediction of nucleosome positioning in *Saccharomyces cerevisiae* were proposed.<sup>9,10</sup> These models are generative and used dinucleotide and pentanucleotide frequencies collected from a training set of aligned nucleosome-bound sequence to predict histone–DNA interactions. Peckham *et al.*<sup>11</sup> proposed a complementary computational model that is discriminative, rather than generative, and that focuses on only sequences that show the strongest signals of nucleosome occupancy or

vacancy in a microarray-based assay of *S. cerevisiae*. Miele *et al.*<sup>12</sup> used sequence-dependent DNA flexibility and an intrinsic curvature to predict the nucleosome occupancy along the genomes of *S. cerevisiae* and *Drosophila melanogaster*. Morozov *et al.*<sup>13</sup> used bending data to design both strong and weak histone-binding sequences, and measured the corresponding free energies of nucleosome formation in order to provide a physical explanation for the intrinsic sequence dependence of histone–DNA interactions. A reference map of nucleosome positions in *S. cerevisiae* has recently been compiled through the analysis of several genome-wide data sets produced in various laboratories and reported at <http://atlas.bx.psu.edu/>, where it has been made available as a Microsoft Excel worksheet named

'additional data file 1'.<sup>14</sup> In yeast, non-random positioning of nucleosomes has been established mainly at promoters, where a nucleosome-free region (NFR) is followed by a highly positioned nucleosome at the +1 position. The +1 nucleosome forms a barrier against which subsequent nucleosomes with a decreasing stability are located.<sup>15</sup> The role of intrinsic DNA sequences either in *in vitro* or *in vivo* nucleosome positioning has recently been at the centre of a debate<sup>16–19</sup> about the existence of a genomic code for nucleosome positioning. One central topic of the debate concerns the correct use of the terms 'positioning' and 'occupancy', which are both connected to the uncertainty on the location of the histone dyad axis on the DNA sequence. In this paper, we adopt the following simple nomenclature: 'positioning' locates the dyad axis with an uncertainty of a few base pairs, whereas 'occupancy' allows reaching a few tens of base pairs, and this error in location is called fuzziness. It is generally accepted that the DNA sequence influences, to some extent, *in vivo* nucleosome positioning, which is regulated by many other additional epigenetic factors. In yeast, the role of chromatin-remodelling enzymes, such as *isw2*, has extensively been studied and a map of genome-wide nucleosome positions in an *isw2* deletion strain has been reported in the 'additional data file 1'.<sup>14</sup> The study of *isw2* has implications for the predictions of nucleosome positions on the basis of the DNA sequence alone and it has been hypothesized that the loss of *isw2* would allow nucleosomes to adopt their inherent positioning preference.<sup>20</sup> In previous papers,<sup>21,22</sup> we reported a correlation between the stability of nucleosomes and the presence of a symmetrical distribution of distances from the nucleosomal dyad axis at specific points. These points represent the locations of phosphates along the DNA minor groove, where histone–DNA electrostatic bonds are formed. These symmetric distances were calculated by means of a tetranucleotide helical rise code<sup>22</sup> applied to known nucleosome-positioning sequences. In that paper, the presence of symmetric distances was mainly correlated to the probability of finding nucleosomes, and on the other hand, we succeeded in observing the absence of symmetric distances in NFRs, which are usually found *in vivo* at the transcription start sites (TSSs) of human promoters. We recently extended our analysis to yeast promoters and observed that helical rise profiles of NFR sequences exhibit lower mean values than those obtained for adjacent sequences where nucleosomes are usually found. In addition, we argued that the generation of mean helical rise curves and the presence of a symmetric shape in their profile could be related to the presence of nucleosomes. These preliminary results prompted us to analyse the whole yeast

genome in a systematic way, so that we generated profiles of helical rise values averaged by various smoothing windows (see methods) and searched for possible correlations between helical rise profiles and known nucleosome positions. We hypothesized that the histone protein begins to interact with DNA exploring at least 26 bp, a length where six principal interactions between the DNA phosphates and the basic amino acids of H3 and H4 histones are formed.<sup>23</sup> Once the optimal site of interaction has been found, DNA continues to wrap around the histone protein and 56 bp are involved in the formation of six additional interactions. When 90 bp are engaged in the complex, about a complete turn of DNA is wrapped, while the final form of the nucleosome is achieved when ~150 bp make one and three quarters of a turn of DNA wrap around the histone protein. We obtained then our helical rise profiles by a set of smoothing windows of 26, 56, 90 and 150 bp, respectively, in order to account for the start and the end of histone–DNA interaction. We also compared our helical rise profiles with those obtained by the dinucleotide frequency method developed by Segal *et al.*<sup>9,10</sup> and the results are available at [http://genie.weizmann.ac.il/software/nucleo\\_prediction.html](http://genie.weizmann.ac.il/software/nucleo_prediction.html).

## 2. Materials and methods

### 2.1. Nucleosome positions

Data referring to nucleosome positioning in *S. cerevisiae* and *D. melanogaster* were collected from reference maps of nucleosome positions reported at <http://atlas.bx.psu.edu/>. Nucleosome positions of *Plasmodium falciparum* were downloaded from plasmoDB ([www.plasmoDB.org](http://www.plasmoDB.org)) and those of human chromosome 20, derived from the experimental data of Schones *et al.*,<sup>24</sup> were downloaded from <http://www.gri.seu.edu.cn/icons>.<sup>25</sup>

Nucleosome positions derived by computational analysis of human chromosome 20 were derived from <http://www.gri.seu.edu.cn/icons>.

### 2.2. DNA sequences

The whole nucleotide sequences of *S. cerevisiae*, *D. melanogaster* and *P. falciparum* were downloaded from the NCBI database ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). The DNA sequence of human chromosome 20 was downloaded from the UCSC genome database, built hg 18 (<http://genome.ucsc.edu>). DNA sequences of 1-kb long (symmetric with respect to the potential dyad axis) were extracted (hg 18 built).

### 2.3. Computational analysis

We used a tetranucleotide helical rise code<sup>22</sup> to 'measure' sequence length in terms of helical rise. Each 1-kb sequence was analysed obtaining a local helical rise profile according to the tetranucleotide code. Four different window sizes of 26, 56, 90 and 150 bp, respectively, were used to get smoothed helical rise profiles, in the following way:

$$p_i^W = \frac{\sum_{j=i-W/2}^{i+W/2} p_j}{W+1},$$

$$i = W/2, \dots, 1000 - (W/2) \text{ and } W$$

$$= 26, 56, 90, 150$$

where  $p_j$  is the plain helical rise profile.

## 3. Results and discussion

### 3.1. Symmetrical helical rise profiles

The DNA sequences were downloaded from the NCBI database ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) and transformed, by our tetranucleotide code, into an array of consecutive helical rise values. In yeast, this primary array was averaged by means of four different smoothing windows and, for each window, 61 110 mapped dyad nucleosome positions, belonging to the reference set included in the additional data file 1,<sup>14</sup> were used to obtain 61 110 profiles of 1-kb-long mean helical rise, each containing 500 bp located upstream of the central dyad axis and 500 bp downstream. For each window, the average of the 61 110 profiles was performed for the reference data set and reported in Fig. 1A as a solid black line. Same approach was used for any of the six other data sets contained in the additional data file 1,<sup>14</sup> returning helical rise profiles (data not shown) closely similar to those generated for the reference set. An analogous procedure was executed for the 54 907 nucleosome positions of an *isw2* deletion strain of *S. cerevisiae*<sup>20</sup> reported in the same additional data file 1.<sup>14</sup>

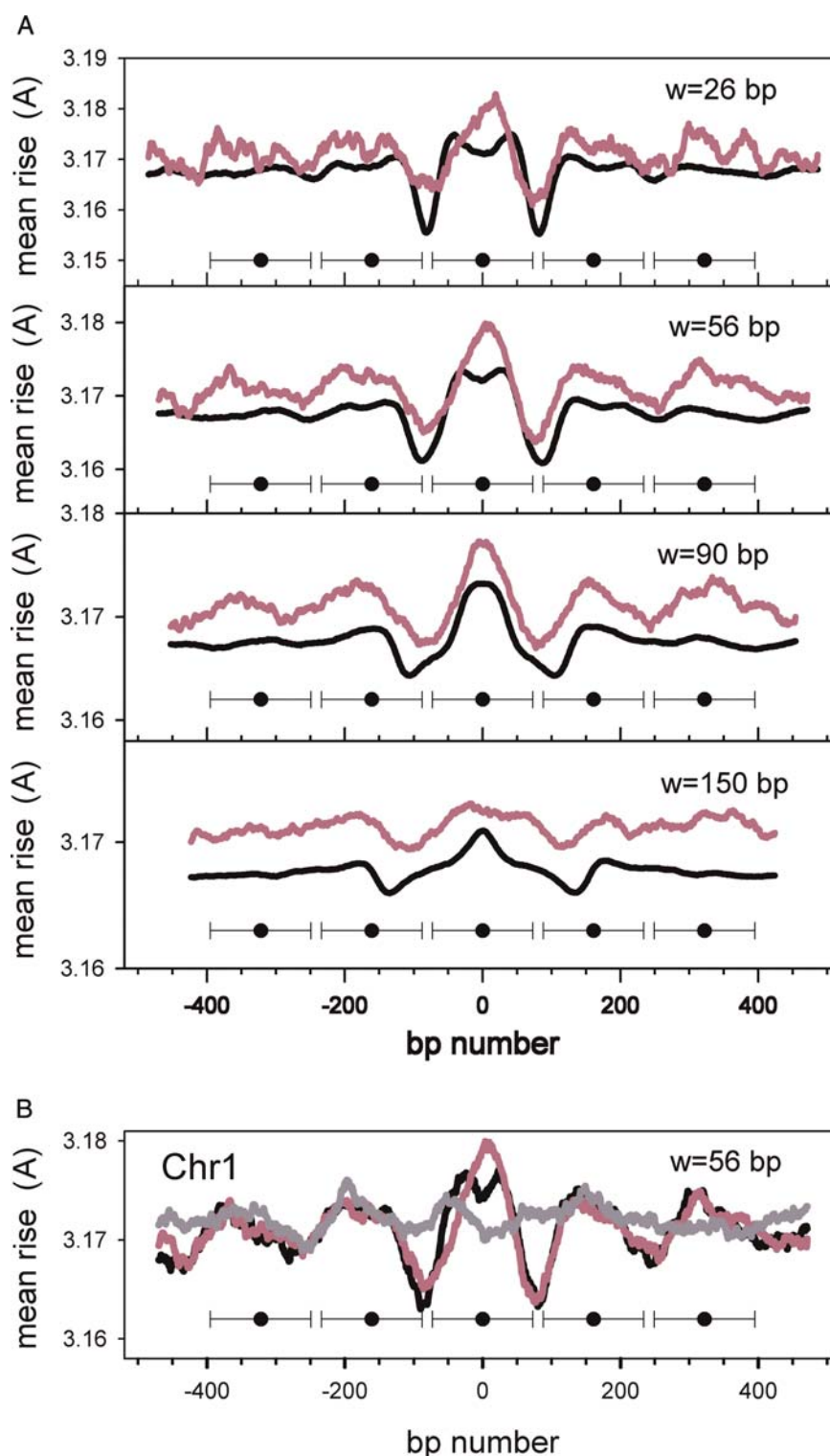
The results obtained for the *isw2* deletion strain are shown in Fig. 1A as a solid pink line. In the same figure are shown, as markers, the positions of five nucleosomes that cover 147 bp each and are mutually separated by 18 bp of linker DNA as reported for yeast.<sup>14</sup> In each of the four panels both the reference and the *isw2* curves are symmetric with respect to the dyad axis located at bp number 0. This behavior is more pronounced for the 90-bp long smoothing window, which represents a length equivalent to a full turn of DNA wrapped around the histone core. When applied to the DNA sequence alone, the helical rise method shows that yeast nucleosomes are positioned

at higher helical rise values than their linkers. The symmetry of the central peak supports our early results<sup>22</sup> on the correlation between nucleosome stability and symmetrical distribution of helical rise values around the nucleosomal dyad axis. In the four panels of Fig. 1A, the mean helical rise values of each *isw2* profile are systematically larger than those of the reference set. Moreover, the reference profiles corresponding to the 26- and the 56-bp-long smoothing windows, unlike their *isw2* counterparts, are both characterized by a central peak clearly split into two adjacent maxima. These maxima are separated by ~60 bp and their disappearance is therefore evident in the helical rise profiles obtained with smoothing windows of 90 and 150 bp reported in Fig. 1A. An explanation for the differences in the rise value and peak shape between the *isw2* and reference samples may be found considering the role of the *isw2*-remodelling enzymes in *S. cerevisiae*. It has been reported that *isw2* enzymes' reposition nucleosomes near promoter regions in order to suppress antisense transcription; the repositioning can range from a few bp to  $\pm 73$  bp and in an *isw2* deletion strain, nucleosomes tend to place themselves at their preferred targets, selected on the basis of the DNA sequence alone.<sup>20</sup> Our algorithm, based on a sequence-dependent approach, correctly shows preferential sites when the effect of *isw2* is silenced. The presence of the double peaks, shown in Fig. 1 for the 26- and the 56-bp-long smoothing windows in the wild-type profiles, is due to the action of *isw2*, which repositions nucleosomes on the DNA sequence. The helical rise can actually render nucleosome occupancy; indeed, in the reference sample, *isw2* remodelers are active and the mapped nucleosome positions are split away from the central position; on the other hand, due to the convergence of the *isw2* nucleosome positions towards a single maximum value in the central peak, the *isw2* average profiles are always higher than those corresponding to the reference set.

Shown in Fig. 1B are the helical rise profiles obtained for the single chromosome I of *S. cerevisiae*. The shape of the black and pink curves (reference and *isw2* data set, respectively) is identical to that reported for the whole yeast genome shown in Fig. 1A at the smoothing window of 56 bp. Also shown in Fig. 1B with a grey continuous line is a control curve obtained for the DNA sequence of chromosome I with dyad positions scrambled with those of chromosome II. The loss of symmetry and the random shape of the scrambled curve are evident.

### 3.2. Peak height and nucleosome occupancy

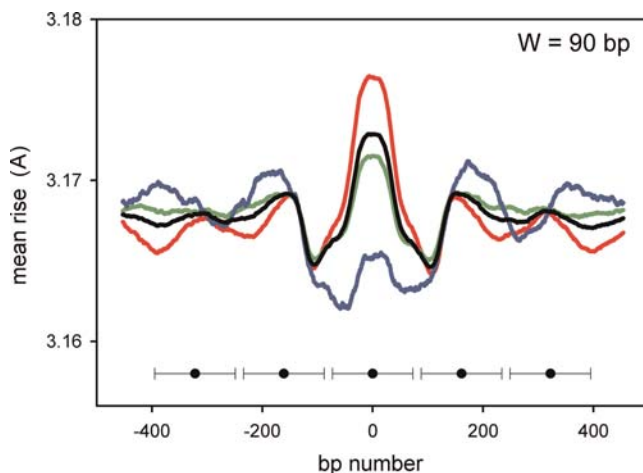
Another characteristic of the profiles shown in Fig. 1 is represented by the lower values of the mean helical



**Figure 1.** Helical rise profiles of nucleosome positions. (A) Average helical rise profiles of mapped nucleosome positions in yeast (average of 61 110 points, black solid line) and in an *isw2* deletion strain (average of 54 907 points, pink solid line) for four different smoothing windows of 26, 56, 90 and 150 bp. (B) Average helical rise profiles of mapped nucleosome positions in chromosome I of yeast (average of 1183 points, black solid line), in an *isw2* deletion strain (average of 1069 points, pink solid line) and in scrambled nucleosome positions in chromosome I of yeast (average of 1183 points, grey solid line) with a smoothing window of 56 bp. Nucleosome dyad positions are aligned at bp number 0; 500 bp upstream and downstream of this point are shown. Black dots with (146-bp long) horizontal bars mark positions of mapped nucleosomes. Colour figures can be found in the online version of this paper.



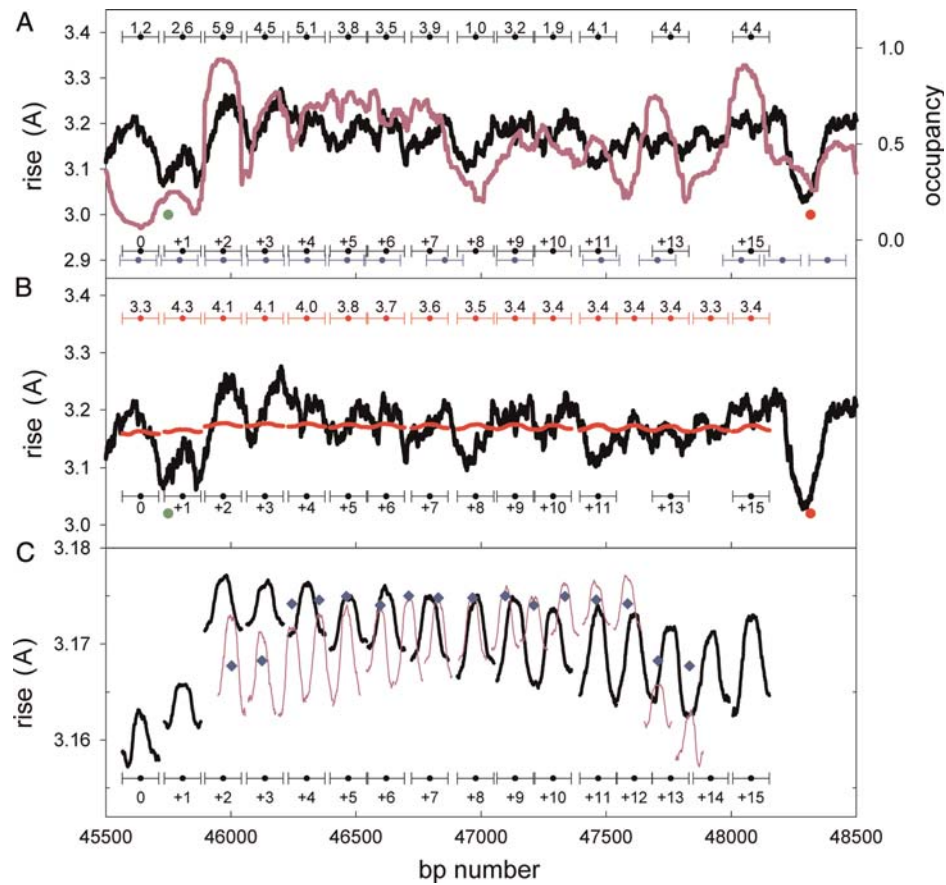
rise for nucleosomes adjacent to the central one. As to this feature, a preliminary attempt was made to understand how the peak height of the helical rise profiles is correlated to nucleosome stability. We reprocessed the reference data set in the additional data file 1,<sup>14</sup> where nucleosome occupancy is expressed in the range from 1 to 7 arbitrary units, classifying nucleosome positions as: very stable (peak height  $>4$ , comprising about 24 600 samples), stable ( $2 < \text{peak height} < 4$ , comprising about 29 800 samples) and weak (peak height  $< 2$ , comprising about 5600 samples). The results obtained for a smoothing window of 90 bp are reported in Fig. 2, where the corresponding overall profile from Fig. 1A is also shown. Data displayed in Fig. 2 state a clear direct relationship between the helical rise peak height and nucleosome occupancy; therefore, our method can be considered to unambiguously assess nucleosome occupancy in a quantitative manner. Furthermore, once again the peaks neighbouring the central maximum exhibit a lower height, except for those marked as weak and represented by a blue solid line. A statistical test (Wilcoxon–Mann–Whitney) was performed between the distribution of helical rise at dyad positions and the distribution of the helical rise at random positions. A  $P$ -value smaller than  $E-16$  assures us that the helical rise is not randomly distributed with respect to nucleosome maps.



**Figure 2.** Peak height analysis of nucleosome positions. Average helical rise profiles of mapped nucleosome positions in yeast for a 90-bp smoothing window. Average of 24 600 points with a peak height  $>4$  (red solid line); of 29 800 points with a peak height  $<4$  and  $>2$  (green solid line); of 5600 point with a peak height  $<2$  (blue solid line); of 61 110 points previously shown in Fig. 1A with a 90-bp smoothing window (black solid line). Black dots with (146-bp long) horizontal bars mark positions of mapped nucleosomes. Colour figures can be found in the online version of this paper.

### 3.3. Helical rise profiles of promoters

The above-mentioned features should reflect the general organization of yeast nucleosomes covering transcribed genes. Figure 3 presents the helical rise profiles of YAL053W, an RNA-polymerase II transcribed gene lying on chromosome 1 whose mapped nucleosome positions are reported in the additional data file 1.<sup>14</sup> In panel A, the helical rise profile of YAL053W smoothed at 90 bp, is shown as a solid black line. Nucleosome positions and peak heights are reported as horizontal bars on the bottom and the top of the figure, respectively. Near the TSS and the transcription termination site (TTS), marked by a green and a red dot, respectively, the rise values reach  $\sim 3.05$  Å, while in the regions where the most stable nucleosomes named +2, +3, +4 etc. are located, the rise values exceed 3.25 Å. The difference of  $\sim 0.2$  Å between the helical rise values of NFRs and those of stable nucleosomes may appear a modest amount, but it must be remembered that the main force involved is of electrostatic nature<sup>26</sup> and follows Coulomb's law, which establishes that the attraction between two charges is proportional to the inverse square of the distance. In addition, the damping down of the attraction depends on the dielectric constant of the medium as well, and free DNA and histone proteins share a hydration shell with a value of dielectric constant equal to 80 due to water. When nucleosomes are formed, the hydration shell gets partially lost and the value of the dielectric constant, although not easy to calculate, is considerably reduced, with a consequent increase in coulombic attraction. The balance between the distance of charges and the stiffness of the hydration shell is likely to be involved in nucleosome formation and mobility. Moreover, it is a known fact that the DNA double helix, when wrapping around a histone in a counterclockwise sense, unwinds and gets stretched,<sup>27</sup> with a consequent variation in the distance between charged phosphate groups along the DNA. In our opinion, the observed preference for higher helical rise values in nucleosome formation is correlated to the lower energetic cost required to stretch DNA in order to form stable interactions between negative DNA phosphates and positive histone groups. In other words, DNA wrapped around a histone protein is constrained and the release of the constraint is better achieved when the mean helical rise is  $>3.2$  Å. In panel A of fig. 3A, the positions of *isw2* nucleosomes are marked by blue horizontal bars. R emodeling of nucleosomes is manifest near NFRs, as well as in regions where the helical rise profile exhibits lower values. This feature is highlighted by the pink profile as well, that shows the occupancy level of YAL053W calculated



**Figure 3.** Helical rise profiles of nucleosome positions at promoters. (A) Helical rise profile (black solid line) of the YAL053W promoter sequence in chromosome 1 of *S. cerevisiae*, obtained by a smoothing window of 90 bp. Nucleosome occupancy profile (pink solid line) obtained by the dinucleotide frequency method by Segal *et al.* Upper black dots with horizontal bars return peak heights of the corresponding nucleosomes. Lower black dots with horizontal bars return nucleosome numbering. Blue dots with horizontal bars return nucleosome repositioning in the *isw2* deletion strain. Green and red dots represent TSS and TTS, respectively. (B) Helical rise profile (black solid line) of YAL053W smoothed at 90 bp as in (A). Average helical rise profiles (red solid lines) for nucleosomes sharing the same position number. Upper red dots with horizontal bars return mean peak heights of the corresponding nucleosomes expressed in arbitrary units as reported in the 'additional data file 1'.<sup>14</sup> Lower black dots with horizontal bars return nucleosome numbering. Green and red dots represent TSSs and TTSs, respectively. (C) Scale enlargement of the mean helical rise profiles for nucleosomes sharing the same position number reported in (B, black solid line). The mirror image of the same profiles (pink solid line). The average between peak heights of the black and pink profiles (blue diamonds). Black dots with horizontal bars return nucleosome numbering. Colour figures can be found in the online version of this paper.

according to the dinucleotide frequency method by Segal *et al.*<sup>9</sup> The black and pink profiles share a few similarities either in the regions from position +2 to +7, where a high value of nucleosome occupancy and of helical rise are found, and in the regions surrounding NFRs like TSS and TTS where low occupancy values high helical rise ones correspond. At positions +8 and +11 a different shape of the helical rise profiles can be observed, characterized by a dyad position at the minimum values of a 'V'-shaped curve instead of the more usual maximum values of a 'peak'-shaped curve. A logical inference is that the central peaks reported in Figs. 1 and 2 are averaged by the sum of both types of helical rise profiles, 'peak'- and 'V'-shaped, so that the average central peaks are always symmetrical but their maximum

values are somehow lowered by the contribution from the 'V'-shaped profile. It must be considered that the major contribution to the mean helical rise profile is due to the peak-shaped curves.

The similarities observed, in Fig. 3, between our curves and those by Segal *et al.* can be explained, considering that nucleosome formation depends on a first step characterized by the bending of DNA and correlated to the dinucleotide distribution, as demonstrated by Segal *et al.*<sup>9</sup> The following step is associated with the stability of bonds formed between DNA and histone proteins, and the second stage is correlated to DNA helical rise.<sup>22</sup>

We analysed the helical rise profiles of all transcribed genes in yeast in order to compare their average peak heights with those reported in Fig. 1A

for the reference data set. In Fig. 3B, the helix rise profile of YAL053W, smoothed at 90 bp, is marked by a black solid line. In the same figure, average profiles for nucleosomes characterized by the same position number are shown. Black horizontal bars mark the nucleosome numbers, whereas red horizontal bars show the average peak height calculated for all the nucleosomes having the same number. It can be seen that the average peak height of the +1 nucleosome is larger (4.3 arbitrary units) than the heights of the following ones, in agreement with the *barrier* model.<sup>15</sup>

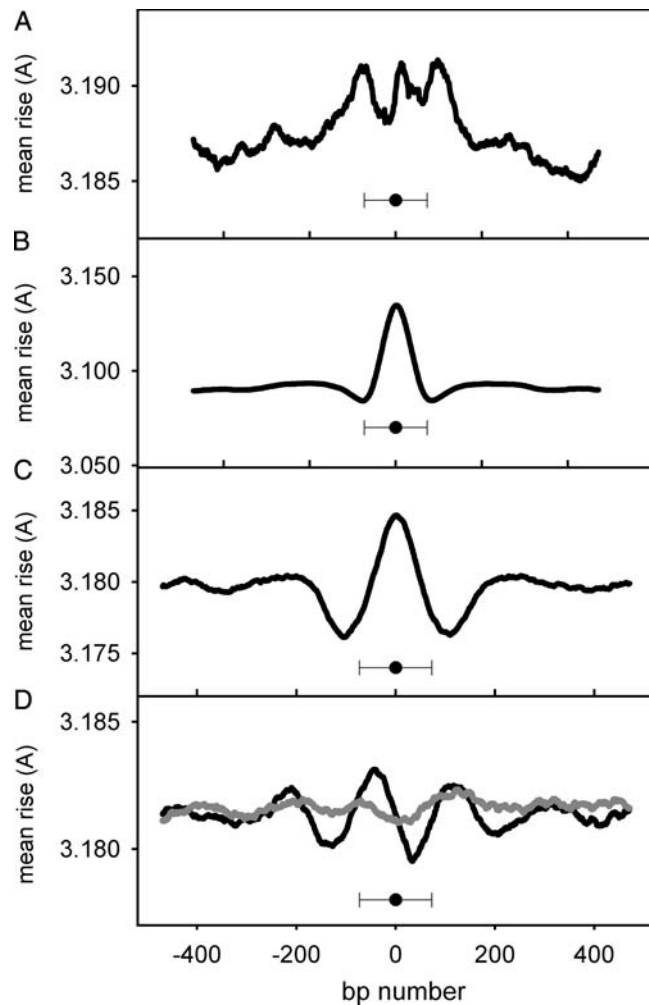
In Fig. 3, average helical rise profiles for nucleosomes sharing the same position number are marked by solid red lines. As apparent from the graph, average helical rise profiles for nucleosomes at positions 0 and +1 are characterized by the lowest values. Statistical tests (non-parametric Wilcoxon–Mann–Whitney) were performed in order to assess the difference between +1 and +2 (and +3, +4, ..., +7) rise value distribution (dyad rise values mediated on a window of 90 bp). *P*-values smaller than  $E-16$  were computed, assessing that the distributions do not belong to the same population, strongly rejecting the null hypothesis. It has recently been reported<sup>16,17</sup> that, in nucleosomes assembled *in vitro*, the usual *barrier-model* pattern, with the +1 nucleosome being the most stable, is not seen. This result supports our idea that the obtained helical rise profiles are mainly correlated to the DNA sequence and can mirror *in vitro* occupancy. In Fig. 3C, average helical rise profiles for nucleosomes sharing the same position number (solid red line in Fig. 3B) are shown on a magnified scale and marked by a solid black line. The presence of a *barrier* beginning at position +2 is now clearer. Note here that the number of nucleosomes used to get the average profiles varies from 7497 (for position +1) to ~1000 (for position +15), reaching the final value of ~48 000 positioned nucleosomes instead of the 61 110 reported in the additional data file 1<sup>14</sup> and used to construct Figs 1 and 2. About half of the 7497 nucleosomes at position +1 reside on opposite strands (3758 on the Watson and 3739 on the Crick strand) and this trend is fulfilled by the following nucleosomes (at increasing position numbers) as well. As a consequence, the global nucleosome pattern must be represented as due to the presence of a mirror image of the original black profiles, as shown in Fig. 3C (pink solid line). Here the blue diamonds represent the average between peak heights belonging to black and pink profiles. The mean helical rise of the blue diamonds ranging from position +2 to position +12 shows an almost constant value of 3.175 Å, similar to the mean rise of the central peak shown in Fig. 1 (black continuous line

with a smoothing window of 90 bp). The mean helical rise of the blue diamonds derived from positions 0, +1, +14 and +13 shows a value of about 3.168 Å similar to the peaks neighbouring the central one in Fig. 1. The adoption of the mirror image in Fig. 3C allowed us to interpret the shape of Fig. 1 mainly due to the contribution of the peaks that constitute the dynamic barrier as shown in Fig. 3C.

### 3.4. Helical rise profiles of additional genomes

Figure 4 presents the helical rise profiles derived from experimentally mapped nucleosome positions in *D. melanogaster*,<sup>28</sup> *P. falciparum*<sup>29</sup> and in chromosome 20 of human genome.<sup>25</sup> All the curves were obtained with smoothing windows of 56 bp in order to assess the presence of peaks split in two adjacent maxima as previously observed in yeast. This pattern is not evident in the helical rise profiles shown in Fig. 4, and only in the case of *D. melanogaster* (Fig. 4A) the presence of three peaks around the central zone of the reference dyad position may be related to the action of remodelling enzymes as previously shown and discussed in Fig. 1 in the case of yeast. The nucleosome positions mapped in yeast have been derived by six different experimental data sets and the best accuracy<sup>20</sup> of tiled microarrays covering the entire yeast genome was extended up to a 5-bp spacing. Only one data set of nucleosome positions is still available for *D. melanogaster*, *P. falciparum* and *Homo sapiens* while the spacing of their tiled microarrays comprises 36, 36 and 25 bp, respectively, therefore the results shown in Fig. 4 must be considered as preliminary ones. A general conclusion is evident from the analysis of the three genomes, i.e. the preferential nucleosome occupancy occurs where the mean helical rise reaches its largest values. Probably the different shape of the three curves reflects the different organization of nucleosome near the +1 position at promoters. In yeast, the +1 nucleosome is usually found at +75 bp from TSS and in *D. melanogaster* this position is shifted at +135.<sup>28</sup> In *Plasmodium*, near promoters are present NFRs that are devoid of nucleosomes;<sup>29</sup> in human expressed and unexpressed genes exhibit differential positioning of nucleosome +1.<sup>24</sup> Figure 4D shows the helical rise profiles derived by nucleosome positions in chromosome 20 of the human genome computationally mapped from the DNA sequence. Dyad positions were elaborated by Liu *et al.*<sup>25</sup> either from their curvature profile method or from curves obtained according to the dinucleotide frequency method of Segal *et al.*<sup>9</sup> A comparison of the helical rise profile reported in Fig. 4C for the experimentally determined dyad positions of human chromosome 20 with the two





**Figure 4.** Helical rise profiles of nucleosome positions of various genomes. Average helical rise profiles of experimentally mapped nucleosome positions in *D. melanogaster* (A, average of 617 304 points, black solid line), *P. falciparum* (B, average of 89 115 points, black solid line), chromosome 20 of human genome (C, average of 65 379 points, black solid line) and in computationally mapped nucleosome positions in chromosome 20 of human genome (D, average of 65 537 points, dinucleotide frequency method, grey solid line and average of 65 537 points, curvature profile method, black solid line). All the curves were obtained with smoothing windows of 56 bp. Nucleosome dyad positions are aligned at bp number 0; 500 bp upstream and downstream of this point are shown. Black dots with (146-bp long) horizontal bars return the position of the central nucleosome.

computationally derived curves of Fig. 4D shows that the predictive power of the curvature profile method (black continuous line) is higher than that attainable with the dinucleotide frequency method (grey continuous line) although for the curve obtained with the curvature profile a remarkable shift of  $\sim 40$  bp in the position of the central peak is observed. We are not able at the moment to explain the reason for this shift.

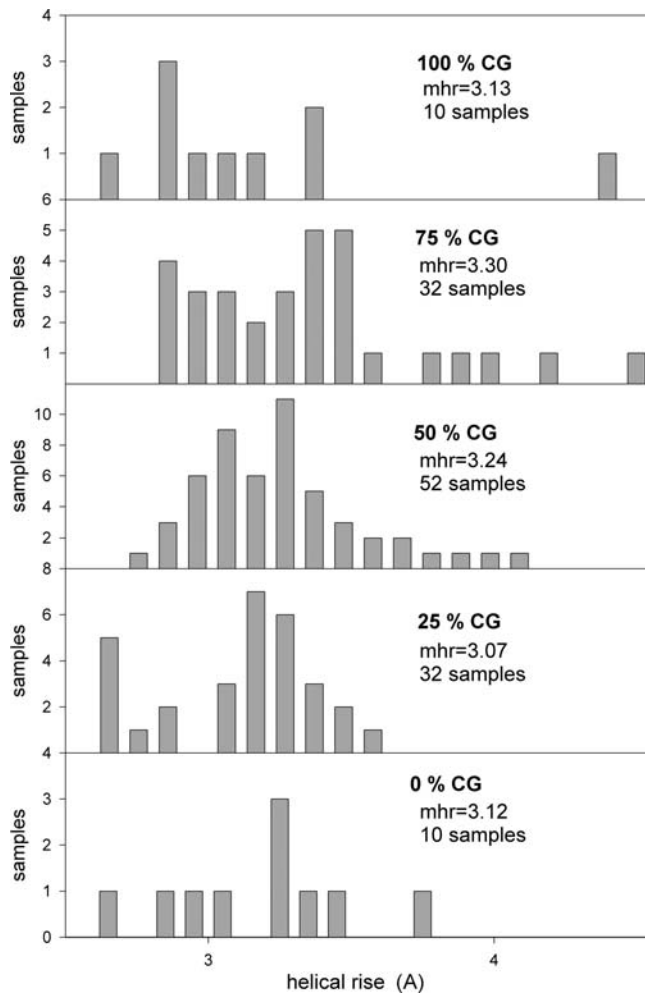
### 3.5. Helical rise and GC content

The GC content of DNA sequences was found to be correlated to the intrinsic nucleosome occupancy in yeast and in *Caenorhabditis elegans*,<sup>30</sup> and a high GC content was found at human regulator sequences enriched in well-positioned nucleosomes.<sup>31</sup> Chen *et al.*<sup>32</sup> reported that in *C. elegans*, the nucleosome-enriched regions are GC-rich when compared with the nucleosome-free sequences and, recently, the GC content has played an important role in the determinants of nucleosome organization in human cells.<sup>33</sup> We have reanalysed the helical rise values of the 136 possible tetranucleotides reported in our previous study<sup>22</sup> and in Fig. 5 we plot the distributions of their values as a function of the five different GC contents. At 100 and 0% GC, the distribution of the helical rise is represented with only 10 samples, the profiles are almost flat and the mean helical rises are similar, 3.13 and 3.12 Å, respectively. The profiles of the 25, 50 and 75% GC show that the mean helical rise increases with the increasing of the GC content ranging from 3.07 to 3.3 Å. The correspondence of high helical rise and GC-rich sequences may constitute a very simple explanation of why our helical rise profiles resemble maps of nucleosome occupancy currently reported in the literature and why nucleosomes are mainly found at positions where helical rise reaches its largest averaged values.

## 4. Conclusions

We used genomic maps of yeast nucleosome positions as primary standards to study the role of DNA helical rise in chromatin organization. Preferential positions for nucleosomes were found where the mean helical rise reaches its largest values at GC-rich DNA sequences. This result was confirmed by a preliminary survey of genomic maps of positioned nucleosomes of the *D. melanogaster* and *P. falciparum* genomes and of the human chromosome 20. We suggest that, in order to relieve the constraint imposed by the bending of the double helix, DNA regions characterized by high helical rise values are favoured when compared with shorter ones, i.e. the former do not need further stretch in order to reach their interaction points and, therefore, the stability of nucleosomes is directly correlated to the mean value of helical rise. In a previous paper,<sup>22</sup> we reported a correlation between the stability of nucleosomes and the presence of a symmetrical distribution of distances from the nucleosomal dyad axis at specific points. The same feature is now observed in yeast helical rise profiles as a series of 'peak'- and 'V'-shaped curves. The difference between the two shapes resides in the lower stability of the 'V'-shaped





**Figure 5.** Helical rise and GC content. The helical rise values of the 136 unique tetranucleotides are shown divided according to their GC content.

profiles, due to their lower helical rise values. In descending order of preference in nucleosome formation, DNA sequences characterized by the smaller helical rise values (i.e. TSS) are placed last.

Preferential nucleosome occupancy was found downstream of NFRs, as observed in the barrier nucleosome model<sup>15</sup> for statistical positioning of nucleosomes in yeast.

### Authors' contributions

F.P. and D.S. equally contributed to the project, both in terms of computational analysis and drafting of the manuscript. All authors read and approved the final manuscript.

### References

1. Lohr, D., Fatchell, K. and Van-Holde, K.E. 1977, On the occurrence of nucleosome phasing in chromatin, *Cell*, **12**, 829–36.

2. Trifonov, E.N. 1980, Structure of DNA in chromatin, *Eur. J. Cell Biol.*, **22**, 74.
3. Trifonov, E.N. and Sussman, J.L. 1980, The pitch of chromatin DNA is reflected in its nucleotide sequence, *Proc. Natl Acad. Sci. USA*, **77**, 3816–20.
4. Mengeritsky, G. and Trifonov, E.N. 1983, Nucleotide sequence-directed mapping of the nucleosome, *Nucleic Acids Res.*, **11**, 3833–51.
5. Wang, Y.H. and Griffith, J.D. 1996, The [(G/C)3NN]n motif: a common DNA repeat that excludes nucleosomes, *Proc. Natl Acad. Sci. USA*, **93**, 8863–7.
6. Lowary, P.T. and Widom, J. 1998, New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning, *J. Mol. Biol.*, **276**, 19–42.
7. Thåström, A., Bingham, L.M. and Widom, J. 2004, Nucleosomal locations of dominant DNA sequence motifs for histone–DNA interactions and nucleosome positioning, *J. Mol. Biol.*, **338**, 695–709.
8. Field, Y., Fondufe-Mittendorf, Y., Moore, I.K., et al. 2009, Gene expression divergence in yeast is coupled to evolution of DNA-encoded nucleosome organization, *Nat. Genet.*, **41**, 438–45.
9. Segal, E., Fondufe-Mittendorf, Y., Chen, L., et al. 2006, A genomic code for nucleosome positioning, *Nature*, **442**, 772–8.
10. Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., et al. 2009, The DNA-encoded nucleosome organization of a eukaryotic genome, *Nature*, **458**, 362–6.
11. Peckham, H.E., Thurman, R.E., Fu, Y., et al. 2007, Nucleosome positioning signals in genomic DNA, *Genome Res.*, **17**, 1170–7.
12. Miele, V., Vaillant, C., d'Aubenton-Carafa, Y., Themes, C. and Grange, T. 2008, DNA physical properties determine nucleosome occupancy from yeast to fly, *Nucleic Acid Res.*, **36**, 3746–56.
13. Morozov, A.V., Fortney, K., Gaykalova, D.A., Studitsky, V.M., Widom, J. and Siggia, E.D. 2009, Using DNA mechanics to predict in vitro nucleosome positions and formation energies, *Nucleic Acids Res.*, **37**, 4707–22.
14. Jiang, C. and Pugh, B.F. 2009, A compiled and systematic reference map of nucleosome positions across the *Saccharomyces cerevisiae* genome, *Genome Biol.*, **10**, R109.
15. Mavrich, T.N., Ioshikhes, I.P., Venters, B.J., et al. 2008, A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome, *Genome Res.*, **18**, 1073–83.
16. Zhang, Y., Moqtaderi, Z., Ratner, B.P., et al. 2009, Intrinsic histone–DNA interactions are not the major determinant of nucleosome positions in vivo, *Nat. Struct. Mol. Biol.*, **16**, 847–52.
17. Zhang, Y., Moqtaderi, Z., Ratner, B.P., et al. 2010, Evidence against a genomic code for nucleosome positioning, *Nat. Struct. Mol. Biol.*, **17**, 920–3.
18. Kaplan, N., Moore, I., Fondufe-Mittendorf, Y., et al. 2010, Nucleosome sequence preferences influence in vivo nucleosome organization, *Nat. Struct. Mol. Biol.*, **17**, 918–20.
19. Kaplan, N., Hughes, T.R., Lieb, J.D., Widom, J. and Segal, E. 2010, Contribution of histone sequence preferences to

- nucleosome organization: proposed definitions and methodology, *Genome Biol.*, **11**, 140.
20. Whitehouse, I., Rando, O.J., Delrow, J. and Tsukiyama, T. 2007, Chromatin remodelling at promoters suppresses antisense transcription, *Nature*, **450**, 1031–5.
  21. Pedone, F., Mazzei, F. and Santoni, D. 2004, Sequence-dependent DNA torsional rigidity: a tetranucleotide code, *Biophys. Chem.*, **112**, 77–88.
  22. Pedone, F. and Santoni, D. 2009, Sequence-dependent DNA helical rise and nucleosome stability, *BMC Mol. Biol.*, **10**, 105.
  23. Richmond, T.J. and Davey, C.A. 2003, The structure of DNA in the nucleosome core, *Nature*, **423**, 145–50.
  24. Schones, D.E., Cui, K., Cuddapah, S., et al. 2008, Dynamic regulation of nucleosome positioning in the human genome, *Cell*, **132**, 887–98.
  25. Liu, H., Duan, X., Yu, S. and Sun, X. 2011, Analysis of nucleosome positioning determined by DNA helix curvature in the human genome, *BMC Genomics* **12**, 72–83.
  26. Ninaber, A. and Goodfellow, J.M. 1999, DNA conformation and dynamics. *Radiat. Environ. Biophys.*, **38**, 23–9.
  27. Yanao, T. and Oshikawa, K. 2008, Elastic origin of chiral selection in DNA wrapping, *Phys. Rev. E*, **77**, 1–11.
  28. Mavrich, T.N., Jiang, C., Ioshikles, I.P., et al. 2008, Nucleosome organization in the *Drosophila* genome, *Nature*, **453**, 358–62.
  29. Ponts, N., Harris, E.Y., Prudhomme, J., et al. 2010, Nucleosome landscape and control of transcription in the human malaria parasite, *Genome Res.*, **20**, 228–38.
  30. Tillo, D. and Hughes, T.R. 2009, G + C content dominates intrinsic nucleosome occupancy, *BMC Bioinformatics*, **10**, 442.
  31. Tillo, D., Kaplan, N., Moore, I.K., et al. 2010, High nucleosome occupancy is encoded at human regulatory sequences, *PLoS ONE*, **5**, e9129.
  32. Chen, K., Wang, L., Yang, M., et al. 2010, Sequence signature of nucleosome positioning in *Caenorhabditis elegans*, *Genomics Proteomics Bioinformatics*, **8**, 92–102.
  33. Valouev, A., Johnson, S.M., Boyd, S.D., Smith, C.L., Fire, A.Z. and Sidow, A. 2011, Determinants of nucleosome organization in primary human cells, *Nature*, **474**, 516–20.